Executive Steering Committee
On A.C.E. Policy II
(ESCAP II)
Report 11
October 10, 2001

ESCAP II: Analysis of Unresolved Codes in Person Matching

Xijian Jim Liu
John A Jones
Roxanne Feldpausch
Decennial Statistical
Studies Division

U S C E N S U S B U R E A U
Helping You Make Informed Decisions

CONTENTS

LIST OF TABLES

EXECUTIVE SUMMARY

This report provides information on the structures of unresolved codes in person matching of the Accuracy and Coverage Evaluation (A.C.E.) and assists the ESCAP committee in reviewing the missing data procedure of the A.C.E.

The unresolved status rates for the 2000 A.C.E. computed in the B series memos are confirmed. These 2000 rates of unresolved codes in person matching were higher than those in 1990, but were not enough to cause major concerns. The distributions and sources of unresolved status were consistent with what we knew and with the missing data procedure we developed.

How did the unresolved status rates in the 2000 A.C.E. compare with those in the 1990 PES?

The **2000 unresolved status rates were higher than those in 1990,** but were not enough to cause major concerns. These unresolved cases were imputed in the missing data procedure and had no direct impact on the total error model.

- In the 2000 A.C.E., 2.2 percent of the P-sample people had unresolved residence status (including 1.2 percent also with unresolved match status), and 2.6 percent of the E-sample people had unresolved enumeration status. The match status and the residence status in the P sample were imputed separately. In 1990 PES, 1.8 percent of the P-sample people had unresolved match status and 1.3 percent of the E-sample people had unresolved enumeration status.

- There was a difference in the coding procedures between 2000 and 1990 that impacted the E-sample unresolved rates. People who did not live at the sample address but had an incomplete Census Day address as indicated on the follow-up form were coded as erroneous enumerations in the 1990 PES and were coded as unresolved status in the 2000 A.C.E. These unresolved cases in 2000 consisted of 0.4 percent of the E sample and were imputed as erroneous enumerations at high probabilities.

- After the before follow-up person matching, certain people were sent for follow-up interviewing to gather more information to determine their match status, residence status or enumeration status. There were 15.6 percent of the P-sample persons in follow-up and 14.6 percent of the E-sample persons in follow-up coded as unresolved status.

How did the characteristics of the unresolved cases in the A.C.E. compare with those in the past?

The distributions and sources of unresolved status were consistent with what we knew and with the missing data procedure we developed.

- Persons collected from proxy initial interviews was a major source of P-sample unresolved status. Among P-sample and E-sample people in the follow-up, a major source of

unresolved status was proxy interviews in the follow-up. A majority of the unresolved cases from follow-up interviews collected partial information. This partial information collected from follow-up interviews was used in the imputation.

- The top reasons for having unresolved status in follow-up interviewing were the reluctance of a respondent to give the Census Day address of a household member who did not live at the address on Census Day and the inability of a proxy respondent to tell whether the person, who lived at the address on Census Day, lived in group quarters or had another residence on Census Day.

- In the P sample, insufficient information for matching and follow-up was the most frequent unresolved code and consisted of 1.2 percent of the P sample (compared with 0.4 percent in 1990). In the E sample, nonmatches with unresolved enumeration status occurred most frequently and consisted of 2.3 percent of the E sample.

What implications do these results have on the adjustment decision?

While the results of the unresolved status did not bear directly on the question, the fact that **things were as we expected** reassures us about the quality of the A.C.E.

1. BACKGROUND

The Accuracy and Coverage Evaluation (A.C.E.) measured the overall and differential coverage of the U.S. population in Census 2000. The A.C.E. comprised of selecting a stratified sample of block clusters independent of the Census, enumerating persons and housing units within these clusters, and matching this enumeration against the Census 2000 enumeration in the corresponding block clusters. The 2000 A.C.E. included an initial housing unit phase, where housing units in the sample block clusters were matched against units listed in the census housing inventory on January 2000; a person interview phase, where demographic information was collected from persons in housing units in the sampled block clusters; and a person phase, where persons in sampled block clusters were matched against the Census record of persons in those same clusters.

The person phase of the A.C.E. began after the initial housing unit phase and the person interview phase had been completed and after P-sample and E-sample housing units had been identified. P-sample housing units were those confirmed to exist within a cluster and chosen for person interviews. E-sample housing units were those listed in the census in the A.C.E. sample block clusters and selected to be in sample after large block subsampling. P-sample persons were defined as Census Day residents in P-sample housing units. They were identified by the interview stage as nonmovers or outmovers or as having unresolved residence status. Furthermore, they must have had a complete or partial person interview. E-sample persons were census persons enumerated in the E-sample housing units.

The person phase consisted of matching P-sample persons against persons listed in the census. Here, the census consisted of E-sample persons and non-E-sample persons who lived in the sampled block clusters. An exception was clusters chosen for Targeted Extended Search. In this case, P-sample persons were matched against census persons both within the block cluster and in blocks that surrounded the cluster.

The person matching consisted of assigning match codes to P-sample persons and census persons. These codes gave information on match status, residence status, enumeration status, and the location of the housing unit. The first element in person matching was a computer match of P-sample persons against census persons within the same block cluster. This match distinguished those persons who were matched from those who were possibly matched and those who did not match. After computer matching occurred, clerks attempted to match all nonmatches and possible matches. After this before follow-up clerical review, certain P-sample and E-sample persons needing additional information to determine their match status, residence status or enumeration status were sent for Person Follow-Up interviews. Information obtained from the follow-up interview was used in the after follow-up coding, and a final match code was assigned. Confirmed nonresidents were removed from the P sample.

Person matching codes indicated when P-sample and E-sample persons were matched, not matched, or possibly matched; they indicated when E-sample persons were correctly enumerated, erroneously enumerated, or the enumeration status was unresolved; and they indicated when P-sample persons were Census Day residents, nonresidents or the residence status was unresolved. The unresolved

codes were the unresolved match status codes, the unresolved enumeration status codes, and the unresolved residence status codes. There was some overlap in these codes; a person can simultaneously had unresolved match and residence status.

The P-sample unresolved match or residence statuses were:

- Persons classified as possible matches by computer matching or during clerical review whose true match status could not be resolved by clerks before the follow-up or in the person follow-up operation (final match code P). The P-sample person also had unresolved residence status.
- P-sample persons identified during computer matching as having insufficient information for matching (final match code KI), and P-sample persons identified during clerical review as having insufficient information for matching because of an incomplete or invalid name (final match code KP). These persons have unresolved match and residence status.
- Matched P-sample persons whose residence status was unresolved (final match code MU).
- Not matched P-sample persons whose follow-up interview could not ascertain their residence status. The probability of residence was imputed for these persons (final match code NU).

The E-sample unresolved match or enumeration statuses were:

- Not matched E-sample persons whose follow-up interview did not reveal enough information to code them as correctly or erroneously enumerated (final match code UE). This match code also pertained to whole household E-sample nonmatches whose geocoding was checked and the interview was unable to determine where the housing unit was located.
- E-sample persons enumerated in housing units identified as in the sample for targeted extended searches (TES), identified as geocoding error and added to the census since January 2000. There was not enough time to do follow-up on these units. Persons residing in these housing units had unresolved enumeration status (final match code GU).
- Matched to P-sample persons with unresolved residence status. The E-sample person had unresolved enumeration status (final match code MU).
- Persons classified as possible matches by computer matching or during clerical review whose true match status could not be resolved by clerks in BFU or in the person follow-up operation (final match code P).

During the missing data procedure, people at addresses where the whole household was coded as insufficient information for matching and follow-up during computer matching (KI) were converted to noninterviews. Other P-sample people with unresolved residence status (KI ,KP, NU, MU, and P) had their probability of residence imputed. P-sample people with unresolved match status (KI, KP, and P) had their probability of match imputed. E-sample people with unresolved enumeration status had their probability of correct enumeration imputed.

This evaluation provides information on the structure of unresolved cases in the A.C.E. and investigates the causes of having unresolved status.

2. METHODS

To investigate the unresolved status in the P sample, we excluded P-sample people coded as confirmed nonresidents because they were removed from the P sample. The percent unresolved residence status was the number of cases having unresolved residence status divided by the total number of P-sample people. The percent unresolved match status was the number of cases having unresolved match status divided by the total number of P-sample people. We used the final P-sample weights that reflected the probability of sample selection in all stages of sampling, and a noninterview adjustment for Census Day interviews.

To investigate the unresolved status among P-sample people in follow-up, we considered all P-sample people in follow-up including confirmed residents, confirmed nonresidents and unresolved residence status. The percent unresolved was the number of cases having unresolved residence status divided by the total number of P-sample people in follow-up. We used the same final P-sample weights mentioned above.

For the E sample, the percent of unresolved enumeration status was the number of cases having unresolved enumeration status divided by the E-sample total. The percent of unresolved enumeration status among E-sample people in the follow-up was the number of cases having unresolved enumeration status divided by the total number of E-sample people in follow-up. We used the final E-sample weights that reflected the probability of sample selection in all stages of sampling.

Standard errors for unresolved rates were produced using stratified Jackknife method and VPLX.

Consider a population of N individuals. Let p_i be the probability of having unresolved status for the ith individual. Assume the probability of having unresolved status can be explained by a logistic model with k independent variables:

$$\log(p_i/(1 ! p_i))=\beta_0+\beta_1 x_{i1}+\beta_k x_{ik}.$$

We used logistic regression to analyze the probability of having unresolved status for people in the P sample and in the E sample. For more on logistic regression, see Hosmer and Lemeshow (1989). We used Taylor linearization method and SUDAAN to compute the Wald Chi-square statistic for all logistic regression analyses.

3. LIMITS

This evaluation was for the fifty states and the District of Columbia only. This evaluation did not consider errors such as

- Errors in coding matching status and residence status,
- Errors in data keying,
- Errors due to nonresponse,
- Response errors,
- Imputation errors,
- Correlation bias.

We determined the interview outcome and respondent type of the follow-up interview using data files keyed from follow-up forms. These data were used for research purposes only. They appeared to be less than perfect, but were the only data available and did not appear to produce significantly discrepant results. The A.C.E. after follow-up coding operation used the actual follow-up forms, not the keyed follow-up data.

4. RESULTS

This section gives the results of this evaluation.

4.1. What were the percentages of P-sample people with unresolved residence status and unresolved match status? What was the percentage of E-sample people with unresolved enumeration status? How did these percentages compare with those in 1990?

Table 1 shows the unresolved status in the P sample for the 2000 A.C.E. and the 1990 Post Enumeration Survey (PES). In the A.C.E., people with unresolved match status also had their residence status unresolved. The match status and residence status were imputed separately in 2000. In 1990, one imputation scheme was used for all P-sample unresolved cases. Although the procedures in 2000 and in 1990 were not identical, one way to compare the unresolved rates was to use the percentages of P sample that had status imputed, i.e., to compare the 2.2% unresolved residence status in the A.C.E. with the 1.8% unresolved match status in the 1990 PES. The difference of 0.4% (standard error=0.09%) is statistically significant at the 0.10 level. This difference could be partly explained by the large increase of insufficient information for matching and follow-up (See 4.2.1).

Table 1. Percent unresolved status in the P sample

P-sample unresolved status	Weighted count	Percent of P sample (standard error)
Unresolved residence status and unresolved match status (2000)	5,844,272	2.2 (0.04)
Unresolved match status only (2000)	3,099,651	1.2 (0.03)
[1]Unresolved match status (1990)	4,258,423	1.8 (0.08)

[1] The 1990 figures are from Cantwell (2001)

Table 2 shows the unresolved enumeration status in the E sample for the 2000 A.C.E. and the 1990 PES. The difference of 1.3% (standard error=0.13) is statistically significant at the 0.10 level. Some differences in the coding procedures between 2000 and 1990 may partly explain the difference in the unresolved rates. If the follow-up form indicated that an E-sample person did not live at the sample address on Census Day but the Census Day address was incomplete, this person was coded as unresolved status in the 2000 A.C.E., but was coded as an erroneous enumeration in the 1990 PES. In 2000, these unresolved cases consisted of 0.4% of the E sample and were imputed as erroneous enumeration at high probabilities (See Table B-2 in Appendix B).

Table 2. Percent unresolved in the E sample

E-sample unresolved status	Weighted count	Percent of E sample (standard error)
Unresolved enumeration status (2000)	6,890,897	2.6 (0.09)
[1]Unresolved enumeration status (1990)	3,205,626	1.3 (0.10)

[1] The 1990 figures is from Childers (2001)

4.2. Which unresolved status code occurred most frequently? What were the characteristics of persons with unresolved status?

Insufficient information for matching and follow-up and non-matches with unresolved enumeration status occurred most frequently. Proxy interviews were a major source of P-sample unresolved cases.

4.2.1. More than one half of the P-sample people with unresolved status had insufficient information for matching and follow-up.

In the 2000 A.C.E., 1.2% of the P sample had insufficient information for matching and follow-up, compared with 0.4% in the 1990 PES. Table 3 gives the P-sample unresolved status by match code for the 2000 A.C.E. Matches and nonmatches with unresolved residence status had their residence status imputed. Possible matches and insufficient information for matching and follow-up had both residence status and match status imputed (with the exception that persons at addresses where the whole household was computer coded as insufficient information for matching and follow-up were converted to noninterviews).

Table 3. Unresolved status in the P sample by match code

Match code	Total	Imputed as		
		Residents, match	Resident, nonmatch	Nonresident
Matched, unresolved residence status	426,629	386,428	0	40,201
Not matched, unresolved residence status	2,317,992	0	1,511,201	806,791
Possibly matched	47,332	38,701	5,788	2,843
Insuff. infor. for matching and follow-up	3,052,319	2,171,647	404,637	476,035
Whole household insuff. Info.	*652,319*			
Partial household insuff. Info.	*2,400,000*			
All unresolved codes	**5,844,272**	**2,596,776**	**1,921,626**	**1,325,870**
Percent of P sample	**2.2**	**1.0**	**0.7**	**0.5**

4.2.2 Almost ninety percent of the E-sample unresolved enumeration status were non-matches with unresolved enumeration status.

Table 4 gives the E-sample unresolved status by match code for the 2000 A.C.E. The unresolved geography code was assigned to unresolved geography cases in the targeted extended search. Other unresolved geography cases were assigned a code of not matched with unresolved enumeration status. These cases with unresolved enumeration status had their enumeration status imputed.

Table 4. Unresolved Status in the E sample by match code

Match code	Total	Imputed as	
		Correct enumeration	Erroneous enumeration
Matched, unresolved enumeration status	356346	338,501	17,845
Not matched, unresolved enumeration status	6,162,271	4,559,773	1,602,498
Possibly matched	39,355	35,436	3,919
Unresolved geography	332,925	305,562	27,363
All unresolved codes	**6,890,897**	**5,239,272**	**1,651,625**
Percent of E sample	**2.6**	**2.0**	**0.6**

4.2.3. Proxy interviews were a major source of unresolved status in the P sample.

Proxy interviews were interviews with nonhousehold members such as neighbors, apartment managers, or real estate agents. Outmovers were residents at the sample address on Census Day who moved out of the sample address before A.C.E. interview day. Information on whole household outmovers was collected from proxy interviews. Table 5 gives the respondent type of A.C.E. person interviews by mover status. Only 5.6 percent of the P-sample people were collected from proxy interviews. Outmovers were a major source of proxy interviews. Forty-seven percent of the people collected from proxy interviews were outmovers.

Table 5. Respondent type of person interview by mover status

Mover status	Household member	Proxy
Nonmovers	96.9%	3.1%
Outmovers	25.0%	75.0%
All P-sample	**94.4%**	**5.6%**

Almost one half of the cases having unresolved residence status came from proxy interviews. Table 6 gives the unresolved residence status in the P-sample by mover status and respondent type. Column (a) shows the percent unresolved, which was the number of unresolved cases in the category divided by the P-sample total in the same category. Column (b) shows the percent of unresolved, which was the number of unresolved cases in the category divided by the number of all unresolved cases in P-sample.

Table 6. Unresolved status by mover status and respondent type

Mover status and respondent type	P-sample percent unresolved (a)	Percent of unresolved (b)
Nonmover, household member	1.2	47.9
Nonmover, proxy	17.3	23.0
Outmover, household member	9.2	3.6
Outmover, proxy	21.5	25.4
All P-sample	**2.2**	**100.0**

Similar to Table 6, Table 7 gives the insufficient information for matching and follow-up in the P sample by mover status and respondent type. Proxy interviews were also a major source of insufficient information for matching and follow-up. More than one half of the insufficient information for matching and follow-up came from proxy interviews. Computer-coded whole household insufficient information were converted to noninterviews and were not represented by the weighted results in Table 7.

Table 7. Insufficient information for matching and follow-up by mover status and respondent type

Mover status and respondent type	P-sample percent insufficient information (a)	Percent of insufficient information (b)
Nonmover, household member	0.5	41.3
Nonmover, proxy	11.1	28.3
Outmover, household member	4.4	3.3
Outmover, proxy	12.0	27.1
All P sample	**1.2**	**100.0**

4.3 What percent of the follow-up persons had unresolved status? What were the characteristics of the persons in follow-up with unresolved status?

There were 15.6 percent of the P-sample people in follow-up and 14.6 percent of the E-sample people in follow-up having unresolved status. Most of these unresolved cases were from follow-up interviews with a proxy respondent.

4.3.1 Among the P-sample persons in follow-up, 15.6% had unresolved status. Among the E-sample persons in follow-up, 14.6% had unresolved status.

After the before follow-up person matching, 6.5 percent of the P-sample people and 16.5 percent of the E-sample people were sent for follow-up interviews to collect more detailed information to correctly code their residence status, match status or enumeration status.

Table 8 shows the percent unresolved in follow-up by various follow-up categories. In the first column, follow-up categories were defined by before follow-up person match status and preliminary housing unit match status. Throughout this section, percent unresolved for P-sample persons in the follow-up was the number of people in follow-up having unresolved residence status divided by the total number of P-sample people in follow-up (confirmed resident, confirmed nonresident and unresolved residence status), and the percent unresolved for E-sample persons was the number of E-sample people in follow-up having unresolved enumeration status divided by the total number of E-sample people in follow-up (confirmed correct enumeration, confirmed erroneous enumeration, and unresolved enumeration status).

Table 8. Percent unresolved in follow-up

Follow-up category	P sample	E sample
Partial household non-match	12.3	13.1
Conflicting household	27.2	28.0
Matches sent to follow-up	15.8	14.6
Possible match	6.8	6.3
Whole household nonmatch, address matched	22.2	18.8
Whole household nonmatch, address did not match or no HU matching	10.9	10.7
All in follow-up	**15.6**	**14.6**

By match code, an overwhelming majority of the unresolved cases in follow-up were coded as not matched with unresolved status. Table 9 gives the percent of follow-up person by unresolved match code.

Table 9. Percent of follow-up persons by unresolved match code

Match code (unresolved)	P sample	E sample
Matched, unresolved residence status	2.3	0.8
Not matched, unresolved residence status	12.7	13.0
Possibly matched	0.2	0.1
Insufficient information for matching and follow-up	0.4	---
Unresolved geography	---	0.7
Total	**15.6**	**14.6**

4.3.2 Proxy respondents in follow-up interviews were a major source of having unresolved status among the persons in follow-up.

Table 10 and Table 11 give the unresolved status by follow-up interview outcomes. Column (a) shows the number of unresolved cases in the row category divided by the number of follow-up persons in the row category. Column (b) shows the number of unresolved cases in the row category divided by the number of all unresolved cases in follow-up. Column (c) shows the number of follow-up persons in the row category divided by the number of all follow-up persons.

The results in Table 10 and Table 11 indicate that more than one half of the unresolved cases in follow-up came from complete interviews with a proxy respondent and that a majority of the unresolved cases in follow-up came from interviews that collected at least partial information. This partial information was used in the residence status imputation and the enumeration status imputation.

Table 10. P-sample unresolved status by follow-up interview outcome

Interview outcome	P-sample Percent unresolved (a)	Percent of unresolved (b)	P-sample percent of follow-up (c)
Complete, household member	5.7	25.1	68.0
Complete, proxy	31.1	52.3	26.2
Noninterview, or respondent type not clear	60.6	22.6	5.8
All in follow-up	**15.6**	**100.0**	**100.0**

Table 11. E-sample unresolved status by follow-up interview outcome

Interview Outcome	E-sample Percent unresolved	Percent of unresolved	E-sample Percent of follow-up
Complete, household member	4.3	19.3	65.0
Complete, proxy	32.7	61.5	27.4
Noninterview, or respondent type not clear	36.9	19.2	7.6
All in follow-up	**14.6**	**100.0**	**100.0**

4.3.3 For persons whose follow-up interview was with a proxy respondent, incomplete information on group quarters or other residence about the person who lived at the address on Census Day was a leading cause of having unresolved status. For persons whose follow-up interview was with a household member, incomplete Census Day address of the person who did not live at the address on Census Day was a leading cause of having unresolved status.

Table 12 shows the distribution of unresolved cases in follow-up by categories based on the answers to the core residence questions on the follow-up form and by respondent type in the follow-up interview. The core residence questions on the follow-up form were:

- A Census Day residence question that asked whether the person lived at the address on Census Day. If the answer was "No", another question that collected the Census Day address followed;
- A group quarters question that asked whether the person lived at group quarters, such as college dorms, nursing homes etc., on Census Day,
- A question about other residence that asked whether the person had another residence on Census Day.

There were four possible answers to each question: "Yes", "No", "Don't know", or "Refuse". The

answer to a question might also be left blank. To better understand the reasons of having unresolved status in person follow-up interviews, the responses to these three questions were investigated. Unresolved cases were classified into the following categories (see Childers and Liu (2001) for more details):

- *Potentially fictitious*: The answers to all three core residence questions were blank. But there were three knowledgeable sources listed in the space provided on the follow-up form who did not know the person. A fictitious code was given if these three sources were valid. Invalid names or notes written on the follow-up forms might have indications that prevented the clerks from coding fictitious.
- *Lived elsewhere on Census Day*: The answer to Census Day residence question was No. To determine if the address where the person lived on Census Day was inside or outside the cluster (or search area), the interviewer was instructed to put the Census Day address on the follow-up form in the space provided. If this address was invalid or incomplete, the person's enumeration or residence status might be unresolved.
- *Census Day residence unresolved*: The answer to the Census Day residence question was "Don't Know" or "Refused".
- *Group quarters or other residence unresolved*: The answer to the Census Day residence question was "Yes". The answer to the group quarters question or the answer to the other residence question was "Don't Know", or "Refused".
- *Others*: This includes other unresolved cases such as cases that the geography section was not completed, that the interviewer was not able to locate three knowledgeable respondents who did not know the person, or that the other residence address was incomplete when the answer to other residence question was "Yes".

Table 12 shows that the distributions of unresolved cases had quite different patterns for different respondent types. We found that the top reasons of having unresolved status in the follow-up interviews were:

- Incomplete Census Day address information when the respondent was a household member who indicated that the person did not live at the address on Census Day;
- Incomplete group quarters or other residence information when the respondent was not a household member who indicated that the person lived at the address on Census Day.

One interpretation of these findings is that the top reasons of having incomplete information in the follow-up were household members' unwillingness to give the address of the person and proxy respondents' inability to know detailed information such as whether the person lived in group quarters or had another residence.

Table 12. Distributions of unresolved status by respondent type in follow-up

Category	A.C.E. person		Census person	
	Household member	Proxy	Household member	Proxy
Potentially fictitious	5.5	8.4	9.1	9.9
Lived elsewhere	51.4	23.6	41.9	17.1
Census day residence unresolved	10.1	6.8	9.1	6.0
Group quarter or other residence unresolved	15.9	55.1	19.7	58.8
Others	17.1	6.1	20.2	8.2
Total	**100.0**	**100.0**	**100.0**	**100.0**

Potentially fictitious people and people who lived elsewhere on Census Day were imputed at a high probability of nonresident or a high probability of erroneous enumeration. Table B-1 and B-2 in Appendix B show the imputation results of these types of unresolved status.

4.4 What variables were associated with unresolved status?

We used logistic regression models as described in section 2 to analyze the probability of having unresolved status. Four different models were considered.

- a logistic regression model for unresolved residence status among P-sample people in the follow-up;
- a logistic regression model for unresolved enumeration status among E-sample people in the follow-up;
- a logistic regression model for unresolved residence status among P-sample people;
- a logistic regression model for unresolved enumeration status among E-sample people;

For P-sample persons in follow-up, we included variables of follow-up interview outcome, mover status, tenure, age group, and census region. Respondent type in the initial person interview, race domain, sex and mode of interview were not significant and were eliminated from the model. A Wald Chi-square of 1,515.2 for follow-up interview outcome showed a very strong effect on the model. Proxy interviews in person follow-up and cases of noninterviews or respondent type not clear were much more likely to have unresolved status than follow-up interviews with a household member. Besides follow-up interview outcome, two variables, age group and region, had moderate effects on having unresolved status. People of age 18-29 and 30-49 were more likely to have unresolved residence status than people in other age groups. People who lived in the northeast region were the least likely to have unresolved residence status. Detailed logistic regression results

for unresolved residence status in the P-sample follow-up are in Table 13.

For E-sample persons in follow-up, we included variables of follow-up interview outcome, tenure, gender, age group, census response method, census region. Race domain was not significant and was eliminated from the model. Again, follow-up interview outcome showed a very strong effect with a Wald Chi-square 1,867.7. Proxy interviews in person follow-up and cases of noninterviews or respondent type not clear were much more likely to have unresolved enumeration status than follow-up interviews with a household member. Tenure and age group also showed moderate effects. Renters were more likely to have unresolved enumeration status than owners. People of age 18-29 and 30-49 were more likely to have unresolved enumeration status than people in other age groups. Detailed logistic regression results for unresolved enumeration status in the E-sample follow-up are in Table 14.

For unresolved residence status in the P sample, we included variables of respondent type in the person interview, mover status, tenure, age, region and mode of person interview. Table 15 gives the result of this logistic regression. Respondent type had a strong effect on the model. This confirms that proxy interview was a major source of having unresolved residence status in the P sample.

For unresolved enumeration status in the E sample, we included variables of tenure, age, sex, region, race domain and census response method. Table 15 gives the result of this logistic regression. Tenure had the strongest effect on the model. Renters were more likely to have unresolved status in the E sample than owners. Age and census response method also showed moderate effects on the model.

Table 13. Logistic regression for unresolved residence status in P-sample follow-up

Variable	Wald Chi-square	d.f.	Odds Ratios	P-value
Follow-up interview outcome	1,515.2	2		0.0000
Complete with proxy interview			6.27	0.0000
Noninterview/respondent type unknown			23.39	0.0000
Complete with household member			1.00	---
Age group	50.0	3		0.0000
1-17			1.14	0.0725
18-29			1.47	0.0000
30-49			1.40	0.0000
50+			1.00	---
Region	33.8	3		0.0000
Northeast			0.58	0.0000
Midwest			0.82	0.0348
South			0.88	0.1488
West			1.00	---
Mover status	20.5	1		0.0000
Outmover			1.20	0.0000
Nonmover			1.00	---
Tenure	15.1	1		0.0001
Renter			1.25	0.0001
Owner			1.00	---

Table 14. Logistic regression for unresolved enumeration status in E-sample follow-up

Variable	Wald Chi-square	d.f.	Predicted Odds Ratio	P-value
Follow-up interview outcome	1,867.69	2		0.0000
Complete with proxy respondent			9.04	0.0000
Noninterview/respondent type unknown			11.91	0.0000
Complete with household member			1.00	---
Tenure	87.52	1		0.0000
Renter			1.66	0.0000
Owner			1.00	---
Age group	92.38	3		0.0000
0-17			1.22	0.0003
18-29			1.63	0.0000
30-49			1.39	0.0000
50+			1.00	---
Sex	20.91	1		0.0000
Male			1.09	0.0000
female			1.00	---
Census response method	17.23	2		0.0002
Enumerator, proxy			1.25	0.0000
Enumerator, non-proxy			1.10	0.1697
Mail			1.00	---
Region	28.79	3		0.0000
Northeast			0.56	0.0000
Midwest			0.79	0.0066
South			0.77	0.0014
West			1.00	---

Table 15. Logistic regression for unresolved residence status in the P sample

Variable	Wald Chi-square	d.f.	Predicted Odds Ratio	P-value
Respondent type	2,548.14	1		0.0000
Proxy			12.27	0.0000
Nonproxy			1.00	---
Mover status	66.09	1		0.0000
Outmover			1.54	0.0000
Nonmover			1.00	---
Age	262.70	3		0.0000
1-17			1.86	0.0000
18-29			1.75	0.0000
30-49			1.31	0.0000
50+			1.00	---
Mode of interviewing	97.65	1		0.0000
Telephone			0.59	0.0000
Person visit			1.00	---
Race domain	99.72	3		0.0000
Hispanic			1.25	0.0001
Non-Hispanic black			1.51	0.0000
Non-Hispanic Asian			1.43	0.0000
All other races			1.00	---
Region	57.70	3		0.0000
Northeast			0.71	0.0000
Midwest			0.72	0.0000
South			0.95	0.3392
West			1.00	---
Tenure	33.60	1		0.0000
Renter			1.25	0.0000
Owner			1.00	---

Table 16. Logistic regression for unresolved enumeration status in the E sample

Variable	Wald Chi-square	d.f.	Predicted Odds Ratio	P-value
Tenure	380.89	1		0.0000
Renter			3.44	0.0000
Owner			1.00	---
Age	203.91	3		0.0000
1-17			1.11	0.0335
18-29			1.89	0.0000
30-49			1.25	0.0000
50+			1.00	---
Census response method	200.14	2		0.0000
Enumerator, proxy			2.30	0.0000
Enumerator, nonproxy			1.64	0.0000
Mail			1.00	---
Sex	67.82	1		0.0000
Male			1.15	0.0000
Female			1.00	---
Region	30.75	3		0.0000
Northeast			0.60	0.0000
Midwest			0.79	0.0175
South			0.99	0.8724
West			1.00	---
Race domain	13.80	3		0.0032
Hispanic			1.14	0.0611
Non-Hispanic black			1.16	0.0132
Non-Hispanic Asian			1.27	0.0045
All other races			1.00	---

5. CONCLUSION

In the 2000 A.C.E., 2.2% of the P sample had unresolved residence statuses (including 1.2% also had unresolved match status), and 2.6% of the E sample had unresolved enumeration status. There were two separate imputations, one for match status and the other for residence status in 2000. In the 1990 PES, 1.8% of the P sample had unresolved match status and 1.3% of the E sample had unresolved enumeration status.

There was a difference in the coding procedures between 2000 and 1990 that impacted the unresolved rates. People who did not live at the sample address on Census Day and had an incomplete Census Day address as indicated on the follow-up form were coded as erroneous enumerations in the 1990 PES but were coded as unresolved status in the 2000 A.C.E. In the 2000 A.C.E., this type of unresolved cases consisted of 0.4% of the E sample and were imputed as erroneous enumerations at high probabilities.

Of the unresolved codes in the P sample, insufficient information for matching and follow-up occurred most frequently and consisted of 1.2% of the P sample (compared with 0.4% in 1990). Of the unresolved codes in the E sample, nonmatches with unresolved enumeration status occurred most frequently and consisted of 2.3% of the E sample.

Almost one half of the P-sample persons having unresolved residence status consisted of persons from proxy interviews. Logistic regression analysis showed that respondent type in person interviewing had strong effects on having unresolved residence status in the P sample. It also showed that age group had a moderate effect on having unresolved residence status in the P sample and that tenure, census response method, and age had moderate effects on having unresolved enumeration status in the E sample.

There were 15.6% of the P-sample people in follow-up coded as having unresolved status, and 14.6% of the E-sample people in follow-up coded as having unresolved status. A leading source of unresolved status in the follow-up was proxy respondents in follow-up interviews. More than one half of the unresolved cases in follow-up were persons whose follow-up interview was a proxy interview. Logistic regression results also indicated that proxy interviews in the follow-up had a strong effect on having unresolved status for both P-sample persons and E-sample persons in the follow-up. A majority of the unresolved cases came from interviews that collected partial information. This partial information was used in the residence status imputation and enumeration status imputation.

The top reasons for having unresolved status in follow-up interviews were the reluctance of a household member to give the Census Day address of the person who did not live at the address on Census Day and the inability of a proxy respondent to tell whether the person, who lived at the address on Census Day, lived in group quarters or had another residence on Census Day.

In summary, the A.C.E. unresolved rates computed in the B series memos (Cantwell (2001)) were confirmed. These 2000 unresolved rates were higher than those in 1990, but were not enough to cause major concerns. The distributions and sources of unresolved cases were consistent with what

we knew from the past and with the missing data procedure we developed.

6. REFERENCES

Cantwell, Patrick J., et al. (2001) "Accuracy and Coverage Evaluation: Missing Data Results", DSSD Census 2000 Procedures and Operations Memorandum Series, B-7.

Childers, Danny R. (2000) "Accuracy and Coverage Evaluation: The Design Document", DSSD Census 2000 Procedures and Operations Memorandum Series, S-DT-1.

Childers, Danny R. (2001) "1990 E-sample Documentation", DSSD Memorandum.

Childers, Danny R. and Liu, Xijian Jim (2001) "Coding of Potentially Fictitious and Lived Elsewhere on Census Day", DSSD Census 2000 Procedures and Operation memorandum Series, S-DT-7.

Haines, Dawn (2001) "Accuracy and Coverage Evaluation Survey: Computer Specifications for Person Dual System Estimation Output Files", DSSD Census 2000 Procedures and Operation memorandum Series, Q-38.

Hogan, Howard (1993) "The 1990 Post-Enumeration Survey: Operations and Results", Journal of the American Statistical Association, Vol 88, No 423., 1047-1060.

Hosmer, David W. and Lemeshow, Stanley (1989) "Applied Logistic Regression", John Wiley & Sons, New York.

Phillps, Dan, (2001), "Hundred Percent Census Unedited File (HCUF)", 2000 Decennial Census Documentation

Raglin, David et al (2001), "Detailed Keying and Output Specifications for the Census 2000 Accuracy & Coverage Evaluation (ACE) Person Followup Form", DSSD Census 2000 Procedures and Operation memorandum Series, S-PT-21.

Appendix A. Percent unresolved by selected variables

Numbers shown in parentheses are standard errors.

Table A-1. Percent unresolved by census region

Region	P-sample	E-sample
Northeast	1.8 (0.09)	1.9 (0.16)
Midwest	1.8 (0.07)	2.1 (0.15)
South	2.6 (0.08)	2.9 (0.15)
West	2.7 (0.11)	3.3 (0.24)

Table A-2. Percent unresolved by regional office

Regional office	P-sample	E-sample
Boston	0.7 (0.07)	1.0 (0.17)
New York	3.0 (0.21)	3.0 (0.42)
Philadelphia	2.6 (0.19)	2.5 (0.28)
Detroit	1.5 (0.10)	1.9 (0.21)
Chicago	1.9 (0.13)	2.3 (0.30)
Kansas City	1.7 (0.12)	2.1 (0.20)
Seattle	2.5 (0.15)	2.8 (0.35)
Charlotte	2.6 (0.13)	3.4 (0.38)
Atlanta	2.8 (0.14)	2.7 (0.19)
Dallas	2.1 (0.13)	2.6 (0.21)
Denver	2.6 (0.18)	3.1 (0.30)
Los Angeles	2.9 (0.20)	3.6 (0.47)

Table A-3. Percent unresolved by type of enumeration area

Type of enumeration area	P-sample	E-sample
Mail out/mail back	2.4 (0.05)	2.8 (0.10)
Update/leave	1.6 (0.08)	1.6 (0.11)
List/enumerate	4.6 (3.66)	14.5(12.45)
Rural update/enumerate	2.5 (0.57)	2.2 (0.50)
Urban update/leave	1.1 (0.43)	0.9 (0.34)
Urban update/enumerate	3.6 (2.38)	1.8 (1.50)
MO/MB to U/L conversion	1.7 (0.57)	0.9 (0.48)

Table A-4. Percent unresolved by tenure

Tenure	P-sample	E-sample
Owner	1.5 (0.04)	1.3 (0.06)
Renter	4.0 (0.10)	5.6 (0.23)

Table A-5. Percent unresolved by MSA

MSA	P-sample	E-sample
Large MSA MO/MB	2.6 (0.09)	2.8 (0.17)
Medium MSA MO/MB	2.4 (0.08)	2.9 (0.15)
Small MSA MO/MB	2.0 (0.08)	2.6 (0.21)
All other TEAs	1.7 (0.09)	1.8 (0.20)

Table A-6. Percent unresolved by return rate

Return rate	P-sample	E-sample
High	2.1(0.05)	2.3(0.10)
Low	2.6(0.10)	3.2(0.19)
Not applicable	2.7(0.16)	3.4(0.28)

Table A-7. Percent unresolved by type of basic address

Type of basic address	P-Sample	E-Sample
Single unit	1.8 (0.04)	1.7 (0.06)
Multi unit	4.1 (0.13)	6.2 (0.33)
Mobile home not in park	2.1 (0.18)	--
Mobile home in park	3.4 (0.39)	--
Single unit in special place	1.9 (0.99)	--
Multi-unit in special place	6.8 (1.30)	--
Others	3.0 (2.83)	--

Table A-8. Percent unresolved by (initial) address code

Address code	P-sample	E-sample
HU matched	1.8 (0.04)	1.6 (0.05)
HU not matched	2.8 (0.22)	8.6 (0.83)
Blank HU match code	---	9.6 (1.36)
Conflicting HH	32.5 (1.21)	25.6 (1.00)

Table A-9. Percent Unresolved by race domain

Race Domain	P-sample	E-sample
American Indian on Reservations	1.6 (0.20)	2.7 (0.93)
American Indian off Reservations	2.6 (0.45)	3.2 (0.45)
Hispanic	2.9 (0.12)	4.0 (0.24)
Non-Hispanic black	3.5 (0.13)	3.8 (0.19)
Hawaiian or Pacific Islander	1.9 (0.40)	3.1 (0.68)
Non-Hispanic Asian	2.8 (0.19)	3.5 (0.33)
Non-Hispanic white	1.9 (0.04)	2.1 (0.90)

Table A-10. Percent unresolved by age/sex group

Age/sex group	P-sample	E-sample
1-17	2.7 (0.07)	2.4 (0.10)
18-29 Male	3.7 (0.12)	5.3 (0.26)
18-29 Female	3.2 (0.12)	4.8 (0.24)
30-49 Male	2.3 (0.07)	2.8 (0.11)
30-49 Female	1.9 (0.06)	2.2 (0.09)
50+ Male	1.4 (0.06)	1.7 (0.08)
50+ Female	1.3 (0.05)	1.4 (0.07)

Table A-11. Percent unresolved by type of return

Type of return	E-sample
Mail	1.9 (0.08)
Enumerator	4.8 (0.21)
Household member	*4.4*
Proxy	*7.1*

Table A-12. Percent unresolved by form type

Type of return	E-sample
Short	2.7 (0.09)
Long	2.2 (0.11)

Table A-13. Percent unresolved by mode of interview

Type of return	P sample
Telephone	0.8 (0.04)
Personal visit	3.1 (0.06)

Appendix B. Imputation Results for Potentially Fictitious People and People Who Lived Elsewhere on the Census Day

Table B-1. Imputation results of unresolved status in the P sample by reason

Reason for having unresolved status	Total	Imputed as		
		Residents, match	Resident, nonmatch	Nonresident
Potentially fictitious	201,233	2,665	27,877	170,691
Lived elsewhere on Census Day	679,713	3,907	95,518	580,289
Other unresolved status	4,963,325	2,590,205	1,798,231	5,74,889
P sample	**5,844,272**	**2,596,776**	**1,921,626**	**1,325,869**

Table B-2. Imputation results of unresolved status in the E sample by reason

Reason for having unresolved status	Total	Imputed as	
		Correct enumeration	Erroneous enumeration
Potentially fictitious	577,910	36,586	541,324
Lived elsewhere on Census Day	1,091,666	245,195	846,472
Other unresolved status	5,221,320	4,957,491	263,829
E sample	**6,890,897**	**5,239,272**	**1,651,625**

Appendix C. Technical Documentation

C.1. A list of files and variables used to produce the results in this report.

C.1.1 P-sample person dual system estimation output file

File name: PDSEUS.DAT (see Haines (2001) for detail file specifications)

MOVERPER.(person mover flag),
REGION (census region),
TEA (type of enumeration area),
TENURE2 (recoded tenure),
TOBA (type of basic address),
DOMAIN (A.C.E. race/Hispanic origin domain),
AGESEX (age/sex post-stratification variable),
RRATEIND (return rate indicator),
MSATEA (MSA/TEA poststratification variable),
AGE2 (Age category),
RPROB (probability of residence),
MPROB (probability of match),
TESFINWT (P-sample final TES-adjusted weight for Census Day),
PROXYIN (proxy/nonproxy respondent),
WPINIT (initial whole/partial match code),
ADDCDE (initial address code),
FUFLAG (follow-up flag),
FU_CODE1 (recoded after follow-up code),
FU_CODE2 (recoded after follow-up code),
BFUMAT (before follow-up person match code),
FINMAT (final match code).

C.1.2. E-sample person dual system estimation output file

File name: EDSEUS.DAT (see Haines (2001) for detail file specifications)

REGION (census region),
TEA (type of enumeration area),
TENURE2 (recoded tenure),
STRCDE (structure code),
DOMAIN (A.C.E. race/Hispanic origin domain),
AGESEX (age/sex post-stratification variable),
RRATEIND (return rate indicator),
MSATEA (MSA/TEA poststratification variable),
AGE2 (Age category),
PSEX (sex code),
CEPROBF (probability of correct enumeration),

TESFINWT (E-sample final TES-adjusted weight for Census Day),
WPINIT (initial whole/partial match code),
ADDCDE (initial address code),
FUFLAG (follow-up flag),
FU_CODE1 (recoded after follow-up code),
FU_CODE2 (recoded after follow-up code),
BFUMAT (before follow-up person match code),
FINMAT (final match code).

C.1.3. Housing unit CAPI interview master file

File name: HUINT.SAS$EBDATA

RECTYPE1 (type of records)
- 1=HU record
- 2=ACE person
- 3=PRX person
- 4=OMP person
- 5=deleted record

STYPE (type of interview)
- C=standard interview
- Q=quality assurance

INTFLG (telephone flag)
- 1=personal visit
- 2=telephone

INTDATE (date of interview)
PRXFLG (proxy interview)
RESPNUM (respondent line number)

C.1.4. Hundred-percent Census Unedited File (HCUF)

See Phillps (2001) for detail file specification.

RCMODE (response collection mode)
PFT (form type)
RHHMEM (respondent household member?)

C.1.5. Keyed follow-up form data files

File names: 3K701xxx.Z03, 3K701xxx.Z04, 3k701xxx.Z05 (see Raglin et al (2001) for detail file specification)

RESPREL (follow-up interview outcome code)

C.2. Unresolved status.

- P-sample unresolved residence status codes are: MU, and NU.
- P-sample unresolved residence status and unresolved match status codes are: P, KI and KP.
- E-sample unresolved enumeration status codes are: MU, UE, P, GU.

C.3. Recoded variables:

- Respondent type (Table 5, 6 and 7).

Proxy	PROXYIN=1 or {PRXYFLG=1 or RESPNUM=99}
Non-proxy	Otherwise

- Follow-up category (Table 8) (Applied to records with FUFLAG=1 or 2)

Matches in follow-up	BFUMAT=M
Possible matches	BFUMAT=P
Conflicting household	ADDCDE=4
Partial household nonmatch	WPINIT=1 and BFUMAT …M,or P
Whole household nonmatch, HU mathed	WPINIT=2 and ADDCDE=1
Whole household nonmatch, HU did not match	WPINIT=2 and ADDCDE=2, or 3

- Follow-up interview outcome (Table 10, 11, 12, 13, 14)

Non-proxy	RESPREL=1, 01,001, 0001etc.
Proxy	RESPREL=2, 02, 002, 0002 etc.
Noninterview, or respondent type not clear	Otherwise

- Unresolved category (Table 12)

Potentially fictitious	FU_CODE1=b
Lived elsewhere	FU_CODE1=e, f, or g
Census Day residence unresolved	FU_CODE1=h, or i
Group quarters or other residence unresolved	FU_CODE1=m, n, or p
Others	Otherwise

- Census response method (Table 14 and 16)

Enumerator, non-proxy	PCMODE=2 and PFT=5,6,17,18 and RHHMEM=0,1
Enumerator, proxy	PCMODE=2 and PFT=5,6,17,18 and RHHMEM=2,3
Mail	otherwise

- A.C.E. mode of interviewing (Table 15)

Personal visit	INTDATE>=06182000 and (either STYPE=C and INTFLG=1 or STYPE=Q and INTFLG=2)
Telephone	otherwise